

VALIDASI KONTEN WEB BERDASARKAN ATURAN ISI TEKS

¹Wulan Agustina K. P

²I Wayan Simri Wicaksana

¹ Jurusan Teknik Informatika Fakultas Teknologi Industri, Universitas Gunadarma
(wulan_akp@studentsite.gunadarma.ac.id)

² Pusat Studi Teknologi Sistem Informasi, Universitas Gunadarma,
(iwayan@staff.gunadarma.ac.id)

ABSTRAK

Saat ini, banyak sekali kompetisi yang bermunculan untuk membuat suatu website yang berkualitas. Tujuan penelitian kami adalah untuk memudahkan pengecekan bentuk ataupun konten suatu website secara otomatis. Terlebih dengan jumlah peserta yang banyak. Subyek penelitiannya adalah program yang dilakukan oleh Universitas Gunadarma kepada mahasiswanya untuk membuat suatu personal web site yang memiliki link ke situs resmi Gunadarma serta memenuhi syarat tertentu lainnya. Untuk dapat mengakses suatu website saat ini yang populer menggunakan web browser berbasis grafik. Permasalahannya untuk pengecekan seperti tujuan diatas harus dilakukan secara manual oleh manusia. Sementara pada penelitian ini, ingin dilakukan pendekatan agar pengecekan dapat dilakukan secara otomatis oleh mesin. Metode evaluasi website yang akan dibahas adalah mengevaluasi isi atau konten dari website dilihat dari sisi objectivity. Yang akan kami ukur disini adalah perbandingan teks pada suatu web pages dengan format HTML standart. Apakah sudah sesuai dengan aturan yang diinginkan. Pendekatan yang digunakan adalah mengambil ide dari sistem crawler dan spider yang digunakan oleh search engine dengan memanfaatkan web browser berbasis teks dikombinasikan dengan pengujian persamaan teks. Kontribusi dari penelitian ini dapat dimanfaatkan secara langsung untuk penilaian dalam suatu kompetisi web, programmer ataupun masyarakat awam yang ingin mengetahui adanya links ke suatu homepage tertentu dari suatu webpages yang ingin kita ketahui.

Kata Kunci: kualitas konten, komparasi teks, web browser

1. PENDAHULUAN

Web browser merupakan sebuah perangkat lunak yang memungkinkan pengguna untuk menampilkan dan berinteraksi dengan text, gambar, and informasi yang ada pada halaman web pada sebuah di World Wide Web (WWW) atau local area network (LAN). Web browser, adalah perangkat lunak yang berfungsi menampilkan dan melakukan interaksi dengan dokumen-

dokumen yang disediakan oleh server web. Penjelajah web yang populer adalah Microsoft Internet Explorer dan Mozilla Firefox. Penjelajah web adalah jenis agen pengguna yang paling sering digunakan. Web sendiri adalah kumpulan jaringan berisi dokumen dan tersambung satu dengan yang lain, yang dikenal sebagai World Wide Web.

Dibawah ini adalah jenis web browser lain menurut Joe Landsberger[4] yang

digunakan oleh user sesuai dengan perkembangan teknologi :

- Opera (www.opera.com) menyebut dirinya sebagai browser tercepat. Browser ini dapat berjalan pada platform windows, Mac, Linux, BeOS dan EPOC. Versi terbaru dari software ini diberikan secara gratis.
- NetCaptor (www.netcaptor.com) sebenarnya bekerja dengan Microsoft Internet Explorer, tetapi menawarkan beberapa fitur tambahan dan fungsionalitas seperti kemampuan untuk melihat beberapa situs secara mudah pada saat yang bersamaan dan memiliki fasilitas auto-refresh. Software ini gratis dan diperuntukkan hanya untuk pengguna windows.
- Lynx (lynx.browser.org) adalah browser yang efisien dengan basis text. Digunakan secara luas di pengaksesan sistem informasi secara lokal. Browser ini berjalan dilingkungan UNIX, VMS, Windows, DOS dan OS/2 EMX.
- iCab (www.icab.de) adalah browser khusus bagi pengguna macintosh. Software ini dikembangkan dari awal memang hanya untuk dikonsumsi oleh para pengguna macintosh. Browser ini terdapat dalam dua versi. Versi gratis dan versi komersial dengan beberapa fitur tambahan.
- NeoPlanet (www.neoplanet.com) adalah browser gratis,

mudah dikustomisasikan, dan memiliki fitur pengiriman email sebaik sebuah download manager.

Semenjak ditemukannya internet sebagai salah satu media komunikasi dan informasi, website merupakan produk teknologi informasi yang terbilang paling populer. Keberadaan website tentu memberikan andil yang cukup besar dalam bidang informasi dan komunikasi, namun terkadang banyak website yang hadir tanpa kualitas yang bagus sehingga tidak dapat menyampaikan informasi yang dibawa. Dengan munculnya beberapa browser dengan fiturnya masing-masing, sudah merupakan suatu tugas bagi para pengembang web agar situs web yang dibuat setidaknya kompatibel apabila dijalankan pada browser yang berbeda. Untuk melakukan pengujian tersebut maka diperlukan adanya testing terhadap halaman web yang dibuat. Testing tersebut meliputi[4] :

- Pengujian dari sudut pandang pengguna dengan menggunakan beberapa browser dan dijalankan pada beberapa sistem operasi
- Pengujian terhadap link-link yang tidak sesuai
- Pengujian terhadap navigasi yang buruk
- Pengujian terhadap kesalahan coding
- Pengujian terhadap dukungan coding tertentu oleh browser

Situs web yang berkualitas akan mampu melewati semua pengujian tersebut tanpa adanya perubahan yang mendasar apalagi sampai mengalami perubahan bentuk halaman.

Saat ini, banyak sekali kompetisi yang bermunculan untuk membuat suatu website yang berkualitas. Hal-hal yang

biasanya dijadikan penilaian oleh pihak penyelenggara yaitu :

- a. Website yang dibuat harus memiliki link, logo, gambar dari pihak penyelenggara.
- b. Model website yang dibuat. Apakah dengan HTML standart atau dengan bantuan CMS (*Content Management System*)?
- c. Model link dari website tersebut. Apakah menggunakan kata "www" pada awal nama website yang dituju atau tidak.

Hal-hal inilah yang menjadi dasar dalam penelitian kami untuk memudahkan pengecekan bentuk ataupun konten suatu website secara otomatis. Terlebih dengan jumlah peserta yang banyak. Untuk dapat mengakses suatu website saat ini populer menggunakan web browser, seperti Internet Explorer(IE), Firefox, Opera dan sebagainya. Permasalahannya dengan web browser berbasis grafik untuk pengecekan seperti tujuan diatas harus dilakukan secara manual oleh manusia. Sementara pada penelitian ini, ingin dilakukan pendekatan agar pengecekan dapat dilakukan secara otomatis oleh mesin. Tujuannya adalah untuk membuat pengecekan menjadi lebih efektif dan efisien terlebih dalam masalah waktu.

2. TINJAUAN PUSTAKA

Web Standards merupakan sebuah standard dalam menyusun kode-kode HTML yang baik dan terstruktur. Kebanyakan website yang ada saat ini disusun dengan kode HTML yang tidak terstruktur, hancur dan berantakan. Karena itu World Wide Web Consortium (W3C) dengan ingin menyuruh para pembuat website diseluruh dunia untuk menstandarkan websitenya.

The World Wide Web Consortium (W3C) merupakan sebuah lembaga konsorsium yang membuat dan terus berobservasi dalam pengembangan teknologi web mencakup XML, HTML dan aplikasi-aplikasi lain yang sering digunakan dalam dunia web. Mereka juga selalu mengeluarkan aturan dan standard supaya siapapun yang membuat dan mengimplementasikan selalu memperhatikan berbagai aspek yang fatal seperti kecocokan dengan perangkat dan browser pengakses, pembaca hingga membuat sebuah website yang dapat berjalan bertahun-tahun karena perubahannya mudah. Walaupun W3C bukan satu-satunya standar dalam pembuatan web, namun W3C merupakan lembaga yang sangat besar *effort*-nya bagi dunia web.

Kualitas sebuah website tentunya tergantung kepada pengujian dan pengawasan terhadap kualitas website itu sendiri, banyak sekali metode untuk mengevaluasi website mulai dari sisi teknologi pembangun, teknologi yang digunakan, isi dari website itu hingga faktor legal yang mendukung website itu.

Mengukur kualitas web dapat disebut juga mengevaluasi web. Pengertian evaluasi menurut para ahli seperti dikemukakan oleh Kaufman dan Thomas (1980) merupakan proses untuk menaksir kualitas dari apa yang sedang berlangsung. Pada evaluasi menuntut adanya kriteria tertentu untuk menentukan kualitas dari apa yang dievaluasi. Beberapa hal yang perlu dilakukan untuk mengukur kualitas sebuah web menurut Adi Sumaryadi[7] antara lain:

- General Appearance, Secara umum kita harus melihat tampilan dari website mulai dari desain, tata letak serta kemudahan navigasi. Desain yang bagus bisa dilihat dari kenyamanan pengunjung seperti warna tidak terlalu

mencolok, nyaman dilihat, kekontrasan juga perlu diperhatikan, website yang bagus tidak terlalu gelap dan juga tidak terlalu terang. Tata letak atau blok-blok konten juga harus kita perhatikan, misalkan informasi penting seharusnya disimpan di atas, menu yang mudah diakses oleh pengunjung dan juga lebar halaman website itu sendiri, web yang bagus sudah memperhitungkan kemungkinan-kemungkinan ukuran resolusi monitor sehingga pengunjung tidak banyak menggunakan scrollbar untuk melihat halaman itu. Kemudahan navigasi juga harus diperhatikan dalam tampilan umum, kecenderungan pengunjung akan dibuat bingung ketika terlalu banyak navigasi yang tidak menentu.

URL yang digunakan mudah atau tidak digunakan. kemudahan url untuk dibaca tidak hanya mempermudah pengunjung untuk mengingat url tadi, melainkan banyak crawler dari mesin pencari yang memanfaatkan kemudahan pembacaan url :

- Ease of Use, kemudahan penggunaan semua elemen yang ada dalam website menjadi salah satu faktor kunci dalam mengavaluasi sebuah halaman web.
- Browser Compatibility, website yang bagus seharusnya sudah dicoba di semua kemungkinan browser yang digunakan. Kompatibilitas terhadap semua browser sangat menentukan kualitas website. Sebuah website yang dibangun harus bagus ketika diakses di Internet Explorer, Opera, Mozilla Firefox, Safari bahkan link sekalipun.
- Purpose, beberapa pertanyaan yang dapat diajukan antara lain :
 - Apakah website mengandung tujuan utama

- Apakah ada keterhubungan antara topic dan topic lainnya
- Apakah satu halaman web di turunkan kedalam halaman-halaman lain yang merupakan subtopiknya
- Apakah website ini mempunyai konten yang bagus?
- Apakah ada keterkaitan antara konten yang ada dengan topic yang dibangun.

- Author, apakah pengelola situs dikenal dengan mudah atau tidak? Kemudahan untuk mendapatkan informasi tentang pengelola atau pembuat situs itu menjadi salah satu faktor utama.

- Link Validation, selain kemudahan untuk menavigasi dan kemudahan pembacaan url, validasi link juga harus diperhatikan. Semua link yang ada harus dapat dikunjungi dengan website, baik link internal dalam sebuah halaman (anchor), link antar halaman hingga link yang berhubungan dengan website orang lain. Semua link yang ada harus menghindari beberapa kriteria seperti dibawah ini :

- HTTP errors, hal yang semacam ini disebabkan oleh kesalahan konfigurasi webserver. Apakah ada link yang menampilkan halaman seperti ini ?
- Not found, apakah link ada yang menghasilkan halaman seperti ini ? pesan seperti ini muncul ketika file yang dituju tidak ada.
- URLs not followed, apakah url yang diberikan tidak bisa dibuka ?
- URLs timed out, apakah ada link yang berat ketika diakses?
- Unreachable URLs, apakah ada link yang hanya sebagian dapat diakses atau tidak sempurna?

- Stability, kesetabilan sebuah website menjadi faktor utama juga dalam mengevaluasi sebuah website.

Menurut Jim Kapoun [3] ada lima petunjuk untuk mengevaluasi web yaitu :

a. Authority

Apakah pengelola situs dikenal dengan mudah atau tidak ? Kemudahan untuk mendapatkan informasi tentang pengelola atau pembuat situs itu menjadi salah satu faktor utama.

a. Currency

Periksa berapa banyak link yang mati dalam halaman web tersebut. Apakah link tersebut diperbaharui terus menerus. Dan apakah informasinya tidak *up to date*.

a. Coverage

Jika halaman web memerlukan perangkat lunak khusus untuk melihat informasi, berapa banyak yang hilang jika kita tidak mempunyai perangkat lunak? Apakah untuk memperoleh informasi bisa didapat dengan cuma-cuma atau harus ada suatu pembayaran. Apakah ada pilihan untuk teks saja, atau *frame* atau ada browser yang diusulkan agar tampilannya lebih baik.

a. Objectivity

Periksa isi dari website tersebut, apakah sudah sesuai dengan tujuan. Apa jenis website dan user berpotensi, ada link luar website yang dapat memberikan informasi.

a. Accuracy

Pastikan bahwa pengarang menyediakan e-mail atau suatu kontak address/phone. Mengetahui perbedaan antara pengarang dan Webmaster.

3. METODE PENELITIAN

Pada penelitian ini, untuk dapat melakukan deteksi terhadap konten web pages kami harus melakukan pengecekan source code dari web pages tersebut. Apakah source code yang ada sudah sesuai dengan aturan yang diinginkan. Salah satu caranya adalah dengan melakukan komparasi string. Metode komparasi string dibedakan menjadi 2 yaitu [5] :

1. Komparasi binary
2. Komparasi teks

Perbedaannya adalah jika kita menggunakan metode komparasi binary maka mesin tidak mengabaikan case sensitive pada huruf ang dibaca. Artinya huruf "A" dengan huruf "a" akan dibaca berbeda atau dianggap tidak sama. Di dalam Microsoft Window, urutan order itu ditentukan berdasarkan halaman kode. Tipikal urutan order biner adalah sebagai berikut :

$$A < B < E < Z < a < b < e < z < \tilde{A} < \hat{E} < \emptyset < \tilde{a} < \hat{e} < \emptyset$$

Sedangkan pada metode berdasarkan teks, mengabaikan case sensitive pada pembacaan huruf. Artinya huruf "A" dengan huruf "a" akan dianggap sama. Hasil komparasi teks berdasarkan pada urutan case sensitive ditentukan oleh sistem local. Sewaktu karakter yang sama diurutkan menggunakan komparasi teks, urutan teks tersebut adalah sebagai berikut :

$$(A=a) < (\tilde{A}=\tilde{a}) < (B=b) < (E=e) < (\hat{E}=\hat{e}) < (Z=z) < (\emptyset=\emptyset)$$

Disini kami akan menggunakan teknik atau metode komparasi string berdasarkan teks. Komparasi teks sekarang menjadi salah satu aplikasi penting yang banyak diterapkan pada berbagai bidang seperti pada teknik kompresi, pengenalan pola, biologi komputasional, web searching dan

data cleaning. Percobaan ini dilakukan dengan menggunakan software python yang terhubung dengan web browser teks yaitu Lynx. Untuk melakukan komparasi teksnya kami menggunakan grep.

Langkah penelitiannya adalah yang pertama kita membaca source code dari web pages yang ingin kita cek menggunakan web browser Lynx dengan bantuan software python. Kemudian source code tersebut kita komparasi menggunakan grep dengan kata atau kalimat yang sudah ditentukan. Outputnya akan langsung terlihat, apakah benar di dalam web pages tersebut sudah memnuhi aturan yang diinginkan.

4. HASIL DAN PEMBAHASAN

Untuk menguji penelitian ini, kami melakukan percobaan terhadap 20 alamat website dari mahasiswa yang dalam pembuatan website diharuskan memenuhi kriteria tertentu. Kriteria tersebut adalah sebagai berikut :

1. Di dalam web pages tersebut harus memiliki logo Universitas Gunadarma, dimana logo tersebut terhubung dengan link situs resmi Gunadarma yaitu www.gunadarma.ac.id.
2. Di dalam web pages tersebut juga harus terhubung dengan situs www.iwayan.powernet.or.id.
3. Untuk kedua kriteria di atas akan dilihat apakah alamat link yang ada dengan menggunakan awalan "www" atau tidak.
4. Web site yang dibuat juga harus menggunakan HTML standart, bukan dengan CMS (Content Management System) misalnya dengan Joomla, Mambo an sebagainya.

Hasil Pengujian ini didapatkan hasil capaian akurasi adalah 90 % dengan membandingkan pengujian secara pendekatan otomasi terhadap manual. Dalam hal ini didapati bahwa masih terdapat kesalahan walaupun prosentasenya sedikit.

Berikut akan ditampilkan tabel yang berisi perbandingan pengecekan antara menggunakan mesin dan manual dengan 20 sample web pages

5. KESIMPULAN DAN SARAN

Dapat disimpulkan bahwa berdasarkan hasil penelitian menunjukan bahwa pengecekan secara mesin atau otomasi juga tidak kalah akurat dibandingkan dengan menggunakan pengecekan secara manual. Waktu yang dibutuhkan pun lebih singkat sehingga lebih efektif dan efisien.

Tentu saja penelitian ini masih jauh dari sempurna. Kelemahan-kelemahan yang dimiliki diantaranya mesin yang dibuat belum dapat mengecek seluruh isi web site tersebut, hanya pada web pagesnya saja. Kemudian mesin ini hanya dapat diterapkan pada web site yang dibuat hanya dengan model HTML standart, bukan dengan bantuan CMS seperti web site yang banyak muncul saat ini. Selain itu mesin pun belum dapat mendeteksi dengan sempurna karena teknik yang dilakukan hanya komparasi teks dimana case sensitive diabaikan dan grep yang dilakukan terkadang tidak sesuai.

Kontribusi dari penelitian ini dapat dimanfaatkan secara langsung untuk penilaian dalam suatu kompetisi web, untuk programmer ataupun masyarakat awam yang ingin mengetahui adanya links ke suatu homepage tertentu dari suatu webpages yang ingin kita ketahui.

Tabel 1. Hasil Pengecekan Dengan Menggunakan Cara Manual

Alamat URL	Logo Gunadarma + www.gunadarma.ac.id	Link www.iwayan.powernet.or.id
http://dega.100webspaces.net	Ada	Tidak ada
http://Hannafitriani.bravehost.com	Tidak ada	Tidak ada
http://Alexandernicko.bravehost.com	Ada	Tidak ada
http://daudzone.indo.ms	Ada	Ada
http://olphanlink.indo.vg	Ada	Tidak ada
http://lovedean.100webspaces.net	Ada linknya saja	Tidak ada
http://astroboy.profil.in	Ada	Ada
http://andikapratama.indo.cc	Hanya logo	Tidak ada
http://thekedolz.100webspaces.net	Ada	Ada
http://alfriano.webs.io	Tidak ada	Tidak ada
http://ranto.profil.cn	Ada	Tidak ada
http://geocities.com/yancsen_boy	Ada	Ada
http://adrianmd.site.io	Ada	Tidak ada
http://geocities.com/goeh_204/style/utama.htm	Ada	Tidak ada
http://freewebs.com/yudya	Hanya logo	Tidak ada
http://freewebs.com/tita87	Hanya logo	Tidak ada
http://freewebs.com/mansyah	Ada	Tidak ada
http://freewebs.com/tian87	Hanya logo	Tidak ada
http://anker.band.io	Hanya logo	Tidak ada
http://komiklovers.profil.bz	Ada	Tidak ada

Tabel 2. Hasil Pengecekan Dengan Menggunakan Mesin

Alamat URL	Logo Gunadarma + www.gunadarma.ac.id	Link www.iwayan.powernet.or.id
http://dega.100webspaces.net	Ada	Tidak ada
http://Hannafitriani.bravehost.com	Tidak ada	Tidak ada
http://Alexandernicko.bravehost.com	Ada	Tidak ada
http://daudzone.indo.ms	Ada	Ada
http://olphanlink.indo.vg	Ada	Tidak ada
http://lovedean.100webspaces.net	Tidak ada	Tidak ada
http://astroboy.profil.in	Ada	Ada
http://andikapratama.indo.cc	Ada	Tidak ada
http://thekedolz.100webspaces.net	Ada	Ada
http://alfriano.webs.io	Tidak ada	Tidak ada
http://ranto.profil.cn	Ada	Tidak ada
http://geocities.com/yancsen_boy	Ada	Ada
http://adrianmd.site.io	Ada	Tidak ada
http://geocities.com/goeh_204/style/utama.htm	Ada	Tidak ada
http://freewebs.com/yudya	Ada	Tidak ada
http://freewebs.com/tita87	Tidak ada	Tidak ada
http://freewebs.com/mansyah	Tidak ada	Tidak ada
http://freewebs.com/tian87	Tidak ada	Tidak ada
http://anker.band.io	Tidak ada	Tidak ada
http://komiklovers.profil.bz	Ada	Tidak ada

6. DAFTAR PUSTAKA

- Dubost, Karl. 20 Mei 2008. My Web site is standard! And yours?..
<http://www.w3.org/QA/2002/04/Web-Quality>
- Engle, Michael. 1 Juni 2008. Evaluating Web Sites: Criteria and Tools.
<http://www.library.cornell.edu/olinuris/ref/research/webeval.html>
- Kapoun, Jim. July/August 1998. "Teaching undergrads WEB evaluation: A guide for library instruction." C&RL News : 522-523.
- Landsberger, Joe. 20 Mei 2008. Evaluating Website Content.
<http://www.studygs.net/evaluate.html>
- NN. 14 Juli 2008. Checking If Two Strings Are Equal. http://www.code-vb.com/visual-basic-source_code/vb-string/checking_if_two_strings_are_equal.html
- Risnandia, Ilyanto N. R2, Marissa A3. 2008. "Analisis Moving Contracting Window Pattern Algorithm (MCWPA) Pada String Comparison".
www.stttelkom.ac.id/staf/FAY/kuliah/DAA/2005
- Sumaryadi, Adi. 29 Mei 2008. Evaluasi Website Menggunakan Seven Checklist dan 10C.
<http://www.adisumaryadi.web.id/index.php?tulisan/detail/11/56/tulisan-56.html>